A note on the use of margins to compare distinguishers

Oscar Reparaz, Benedikt Gierlichs, and Ingrid Verbauwhede

KU Leuven Dept. Electrical Engineering-ESAT/COSIC and iMinds
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
firstname.lastname@esat.kuleuven.be

Abstract. Relative distinguishing margins are becoming a popular measure for comparing distinguishers. This paper presents some examples that show that this measure, although informative and intuitively sound, should not be taken alone as benchmark of distinguishers.

1 Introduction

Since the introduction of Differential Power Analysis (DPA) in [3], several different statistical tools called distinguishers have been proposed. Some distinguishers claim to be more efficient assuming a leakage model (like CPA [1]) or more generic (MIA [2] and KS [13]). A recurring topic in the literature is the need for establishing fair criteria to compare distinguishers and extract broad conclusions, more generally applicable than the comparison of outcomes in specific empirical experiments.

The notion of success rate is of extended use to evaluate distinguishers, probably due to the accessible interpretation of the measure. One of the first works theoretically analyzing the behavior of several univariate distinguishers is presented by Mangard et al. in [7]. They show that the (asymptotic) efficiency, measured as the success rate, of distinguishers based on the correlation coefficient, difference of means and Gaussian templates are essentially the same, given the exact (single-bit) model of the power consumption. This result, however, does not generalize to higher-order attacks as shown by Standaert et al. in [10]. Said work shows that in the context of attacking masked implementations, the choice of distinguisher indeed highly affects the success rate achieved in the attack.

However, measures other than the success rate have also been proposed in previous works. Most notably, Whitnall and Oswald formalized the concept of theoretical margins for a distinguisher in [11,12]. This measure provides an improvement and generalization of several other measures [4,5], and it was shown to be more expressive and informative than the success rate [11]. In short, the relative margin measures to what extent the distinguisher value for the correct key hypothesis stands out over other competing distinguisher values, in a normalized fashion.

In the same series of papers [8,11,12], Whitnall et al. introduce a very interesting idea towards separating the intrinsic distinguishing power of a distinguisher

from estimation inaccuracies, both of which affect the success rate. To isolate these two aspects, the distinguisher values are not estimated but directly computed from the probability densities of the simulated leakage via numerical integration. In this approach, the estimation problem (which for some distinguishers is notoriously hard) is worked around. Whitnall et al. apply this technique to theoretically compare several distinguishers and draw the conclusion that MIA and KSA distinguishers have theoretical advantages over CPA and that the underperformance of MIA-like attacks frequently observed in practice is due to estimation errors.

Theoretical margins are receiving an increasing adoption. Recent works have proposed new distinguishers and justified somehow their superiority based on theoretical margin measures [6,13].

Contribution. This paper presents simple counterexamples of distinguishers that exhibit the exact same success rate, yet their theoretical margins' values can be almost arbitrarily different. Hence, theoretical margins should not be used as the sole measure to compare distinguishers.

Notation. A distinguisher is the statistical tool that is used to compare measurements T to key-dependent predictions Z_k in a standard DPA attack. The *distinguisher vector* $D(k)$ is a vector containing distinguisher values for each subkey k. In the simulations of this paper we assume that the leakage T consists of the Hamming weight of the first DES Sbox output Z with additive Gaussian noise, that is $T = \mathrm{HW}(Z_k) + \epsilon$ with $Z_k = \text{DES-Sbox1}(p \oplus k)$. The signal-to-noise ratio (SNR) is defined as $\frac{\mathrm{var}\ [\mathrm{HW}(Z_k)]}{\mathrm{var}\ [\epsilon]}$.

Organization. In Section 2 we present the main idea: several distinguishers are proposed that serve our purpose of showing that taking only the margins into account can lead to misjudgment. In Section 3 we study the behavior of the distinguishers when noise is present.

2 Two distinguishers

In this section, we present two distinguishers D_1 and D_2 that by construction behave exactly in the same way in practice. That is, the two distinguishers will rank key candidates in exactly the same way: the attack using D_1 will be exactly as successful as the attack using D_2. However, the relative and absolute margins for D_1 and D_2 are different.

2.1 Description

The first distinguisher D_1 is the absolute value of Kocher et al. single bit DPA between measurements T and key-dependent predictions Z_k. That is, for each hypothesis k of the key, the distinguisher computes

$$D_1(k) = \left| \widehat{\mathbf{E}}(T|L(Z_k) = 1) - \widehat{\mathbf{E}}(T|L(Z_k) = 0) \right| \tag{1}$$

where L is a function that extracts one bit from the predictions and $\widehat{\mathbf{E}}$ is the sample mean operator. The second distinguisher D_2 is based on D_1. It computes the squared version of D_1 as

$$D_2(k) = [D_1(k)]^2 \tag{2}$$

$$= \left|\widehat{\mathbf{E}}(T|L(Z_k) = 1) - \widehat{\mathbf{E}}(T|L(Z_k) = 0)\right|^2. \tag{3}$$

2.2 Properties

It is not hard to see that D_1 and D_2 are in essence the same distinguisher. For any two key hypothesis, D_1 will rank them in the same way as D_2. This means that an attack using D_1 will be exactly as successful as one using D_2. One can see D_2 as the composition of first computing D_1 and then *squaring* every distinguisher value (i.e., applying the map $x \mapsto x^2$), as Figure 1 (left) shows. Since the map $x \mapsto x^2$ is strictly increasing in $x \geq 0$ (possible values of D_1 will be always $D_1 \geq 0$), it follows from the definition that the order (key ranking) will be preserved. However, as we will see in the next section, D_1 and D_2 have different theoretical relative margins. (

2.3 Margins for D_1 and D_2

For a given distinguisher that produces the distinguishing vector D, the *relative distinguishing margin*[1] is defined as

$$\text{RelMargin}(D) = \frac{D(k^*) - \max\left[D(k)|k \neq k^*\right]}{\text{std}(D)} \tag{4}$$

where k^* is the correct key and std is the sample standard deviation. The sign of this measure indicates whether an attack using the given distinguisher and a "large enough" number of traces would be successful (or not), and the magnitude of the measure, up to what extend the attack was successful (or not.) In what follows, we computed all relative margins by numerical integration as suggested in [12].

We computed the theoretical relative distinguishing margin for D_1 and D_2 and got, respectively, 0.250 and 0.5176 in a noiseless scenario. Both are positive, which means that the attacks would be successful, given enough traces. The fact that the two magnitudes are different means that the theoretical relative distinguishing margin is, in this situation, measuring something that does not relate to the intrinsic distinguishing ability of D_1 or D_2, since it is clear that by construction both distinguishers behave identically.

We push further our study by introducing another pair of distinguishers D_1^{MIA} and D_2^{MIA}. The distinguisher D_1^{MIA} is MIA and is defined as

$$D_1^{\text{MIA}} = I(T; L'(Z_k)) \tag{5}$$

[1] We note that the distinction between *theoretical* distinguishing margins and distinguishing margins is orthogonal to the observations in this paper, and the consequences affect both.

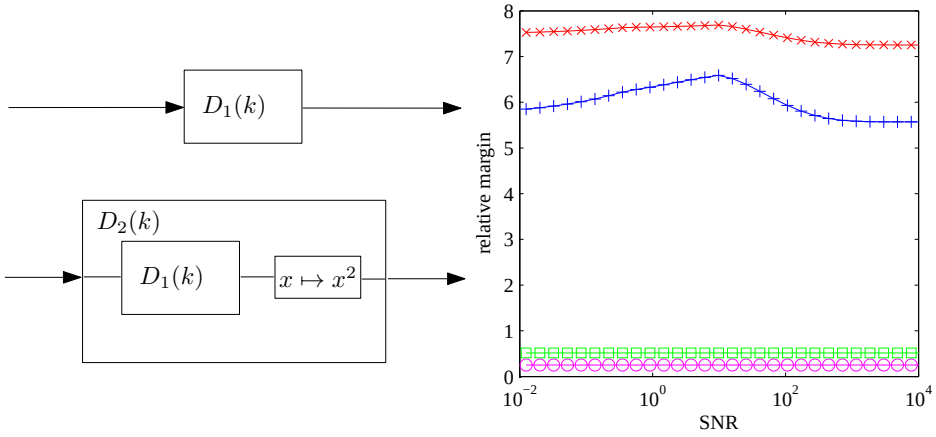

Fig. 1. Left: construction of D_1 and D_2. Right: relative distinguishing margins for D_1 and D_2. Pink $\circ$: margins for D_1. Green $\square$: margins for D_2. Blue $+$: margins for D_1^{MIA}. Red x: margins for D_2^{MIA}.

where $I(\cdot;\cdot)$ denotes Mutual Information and L' is some leakage model. Analogously, we define D_2^{MIA} as the squared version of D_1^{MIA}:

$$D_2^{\mathrm{MIA}} = \left[D_1^{\mathrm{MIA}}(k)\right]^2 \tag{6}$$

$$= \left|I(T; L'(Z_k))\right|^2 . \tag{7}$$

We computed theoretical margins for D_1, D_2, D_1^{MIA} and D_2^{MIA} as a function of the SNR and plot them in Figure 1 (right.) We note that the results of the margin of D_1^{MIA} coincide with those from [12][2]. As expected, the margins for D_1 and D_2 stay constant as the SNR progresses. For the difference of means based distinguishers, noise affects every distinguisher value in the same way, keeping the theoretical distinguishing ability unaffected. For the distinguishers based on MIA the situation is different: margins for D_1^{MIA} and D_2^{MIA} vary as the SNR changes, as [12] pointed out.

From the observation of Figure 1 it is clear that all distinguishers D_1, D_2, D_1^{MIA} and D_2^{MIA} have distinct margins, albeit D_1 (respectively D_1^{MIA}) is essentially the same as D_2 (respectively D_2^{MIA}). Thus, we see that margins do not necessarily relate to success rate. We would incur a misjudgment if based on Figure 1 and without any more information we assess that distinguisher D_2 has more intrinsic distinguishing abilities than D_1. Furthermore, by the same reasoning, from the observation of Figure 1 there is not enough information to claim that distinguisher D_2^{MIA} has more intrinsic distinguishing abilities than D_2, which is a different distinguisher not based on MIA. In the next section we elaborate on the applicability of margins to compare distinguishers.

Note that the observation regarding different margins for D_1^{MIA} and D_2^{MIA} will hold in a theoretical scenario (where there are no estimation errors) as well

[2] Up to a typo in the caption of Figure 2 in [12].

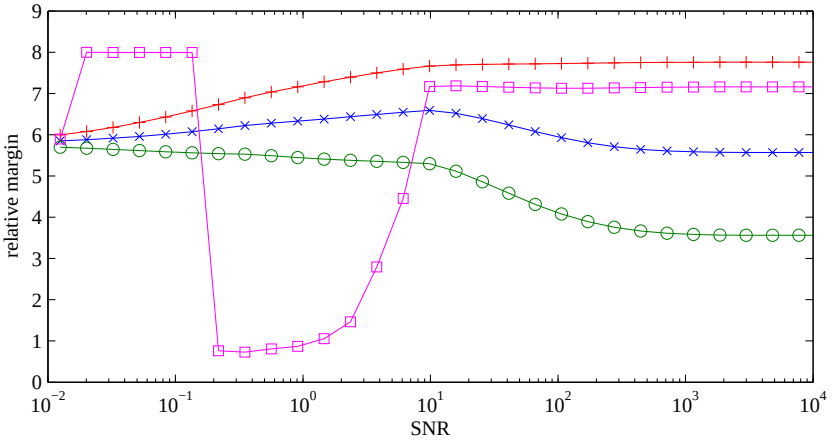

Fig. 2. Red, blue, green: margins for $D^{\mathrm{MIA}}_{a,b}$ for several choices: red, '+': $a = 1.9$, $b = 7$; blue, 'x': $a = 0$, $b = 1$ (this means that $D_{0,1} = D_1$); green, 'o': $a = 0.3$, $b = 0.003$. Pink, □: margins for $D_{f(x)}$ with $f(x) = e^x$ if $x < 0.05$ and $f(x) = 10e^{x+1}$ otherwise ($f(x)$ is strictly increasing in $x > 0$.

as in a practical scenario (since the estimation errors will affect D^{MIA}_1 and D^{MIA}_2 in exactly the same way).

3 Discussion

3.1 The shape of the margins is also different

Upon the observation of Section 2.3, one might ask if the properties of the margin are the same for D^{MIA}_1 and D^{MIA}_2 as the SNR varies . In other words, whether the relative margin for D^{MIA}_2 is just a scaled version of D^{MIA}_1. In this section, we answer this question negatively.

We slightly generalize the construction of D^{MIA}_2. We consider the family of distinguishers $D^{\mathrm{MIA}}_{a,b}$. This family is constructed akin to D^{MIA}_2 but substituting the squaring mapping $x \mapsto x^2$ with a different strictly increasing non-linear mapping $x \mapsto (x + a)^b - a^b$ for some $a, b > 0$. Since the mapping is still strictly increasing in $x \geq 0$, all the distinguishers in the family are essentially the same. We further generalize and also consider the family of distinguishers $D^{\mathrm{MIA}}_{f(x)}$ that is constructed similarly to D^{MIA}_2 but with a generic strictly increasing non-linear mapping $x \mapsto f(x)$. We note that linear mappings of the form $x \mapsto a \cdot x + b$ would not modify relative margins (and will of course lead to attacks with identical success rates.)

In Figure 2 we plot the theoretical relative distinguishing margin for some members of the family of distinguishers previously defined. We can see that the evolution of the relative margin as a function of the SNR can be almost arbitrary, even though all the distinguishers in the figure are essentially the same (they

relate to the same distinguisher up to a strictly increasing non-linear mapping at their output). Thus, one should also be skeptical about drawing conclusions about the behavior of a specific distinguisher from the observation of the shape of the relative distinguishing margin as the SNR varies. In Figure 2, one could assert from the curve corresponding to $D_{0,1}^{\mathrm{MIA}}$ (blue, 'x') that there is a stochastic-resonance-like effect around SNR=10 since the margin achieves a maximum. We note that the very same effect does not exhibit itself for the other equivalent distinguisher in the figure (red, '+'; and green, 'o'.) Therefore, margins alone should not be used to assess the properties of a distinguisher as the SNR varies: distinguisher-specific properties may or may not show in the margins.

3.2 Objection: D_2 is pathologic

One could argue that the construction of appending a non-linear mapping at the output of a previously proposed distinguisher is pathologic. Although D_2 (and subsequent generalizations) was specifically crafted to show the point in this paper regarding relative distinguishing margins and no reasonable person would think that it is any better (or worse) than D_1, we remark that the derived distinguishers are as sound as the original ones. For example, D_2 is as sound as D_1 and still gives a measure of the degree of the correlation between random variables (only in a different scale than D_1), and is as precise as D_1.

3.3 What is left to compare distinguishers?

The task of comparing in a fair way several distinguishers that work on different scales seems hard. One could resort to the well-known success-rate metric, albeit one should be aware of its limitations. Namely, success rates are highly dependent on the statistical estimator used in the computation of the distinguisher values. Besides, once the signal-to-noise ratio is high enough so that the distinguishers under study behave well (they output the correct key hypothesis with high probability, i.e., their success rates reach values close to 1), it becomes hard to compare distinguishers and rank which one is better, since their success rates are all close to 1. On the bright side, success rates are easily computable in empirical settings and can be used to compare distinguishers that work on different scales. The same observations apply to other metrics that are only sensitive to the ordering of the distinguishing vector, such as guessing entropy [9].

4 Conclusion

We showed in this paper that the theoretical relative distinguishing margin can be a useful measure but is not to be used as the sole measure to compare distinguishers, and to assess properties of a specific distinguisher. Although the measure is intuitively useful, and in many cases it informs of useful properties of distinguishers, there are some counterexamples/corner cases shown in this paper where the measure should not be taken solely to judge the behavior of a distinguisher.

Acknowledgments. We thank the anonymous reviewers for their insightful comments. This work was supported in part by the Research Council of KU Leuven: GOA TENSE (GOA/11/007), by the Flemish Government FWO G.0550.12N and by the Hercules Foundation AKUL/11/19. Oscar Reparaz is funded by a PhD Fellowship of the Fund for Scientific Research - Flanders (FWO). Benedikt Gierlichs is Postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO).

References

1. E. Brier, C. Clavier, and F. Olivier. Correlation power analysis with a leakage model. In M. Joye and J.-J. Quisquater, editors, *CHES 2004*, volume 3156 of *LNCS*, pages 16–29. Springer, 2004.
2. B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel. Mutual information analysis. In E. Oswald and P. Rohatgi, editors, *CHES 2008*, volume 5154 of *LNCS*, pages 426–442. Springer, 2008.
3. P. C. Kocher, J. Jaffe, and B. Jun. Differential power analysis. In M. J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397. Springer, 1999.
4. T.-H. Le, J. Clédière, C. Canovas, B. Robisson, C. Servière, and J.-L. Lacoume. A proposition for correlation power analysis enhancement. In *Proceedings of the 8th International Conference on Cryptographic Hardware and Embedded Systems*, CHES'06, pages 174–186, Berlin, Heidelberg, 2006. Springer-Verlag.
5. T.-H. Le, J. Clédière, C. Servière, and J.-L. Lacoume. Noise reduction in side channel attack using fourth-order cumulant. *IEEE Transactions on Information Forensics and Security*, 2(4):710–720, 2007.
6. H. Maghrebi, S. Guilley, O. Rioul, and J.-L. Danger. Some results about the distinction of side-channel distinguishers based on distributions. In *10th International Workshop on Cryptographic Architectures Embedded in Reconfigurable Devices (CryptArchi 2012)*, Saint-Etienne, France, June 19-22, 2012.
7. S. Mangard, E. Oswald, and F.-X. Standaert. One for all - all for one: unifying standard differential power analysis attacks. *IET Information Security*, 5(2):100–110, 2011.
8. E. Oswald, L. Mather, and C. Whitnall. Choosing distinguishers for differential power analysis attacks. In *Non-Invasive Attack Testing Workshop*. NIST, 2011.
9. F.-X. Standaert, T. G. Malkin, and M. Yung. A unified framework for the analysis of side-channel key recovery attacks. In *Proceedings of the 28th Annual International Conference on Advances in Cryptology: The Theory and Applications of Cryptographic Techniques*, EUROCRYPT '09, pages 443–461, Berlin, Heidelberg, 2009. Springer-Verlag.
10. F.-X. Standaert, N. Veyrat-Charvillon, E. Oswald, B. Gierlichs, M. Medwed, M. Kasper, and S. Mangard. The world is not enough: Another look on second-order DPA. In M. Abe, editor, *ASIACRYPT 2010*, volume 6477 of *LNCS*, pages 112–129. Springer, 2010.
11. C. Whitnall and E. Oswald. A comprehensive evaluation of mutual information analysis using a fair evaluation framework. In P. Rogaway, editor, *CRYPTO 2011*, volume 6841 of *LNCS*, pages 316–334. Springer, 2011.
12. C. Whitnall and E. Oswald. A fair evaluation framework for comparing side-channel distinguishers. *Journal of Cryptographic Engineering*, 1(2):145–160, August 2011.

13. C. Whitnall, E. Oswald, and L. Mather. An exploration of the Kolmogorov-Smirnov test as a competitor to Mutual Information Analysis. In E. Prouff, editor, *CARDIS*, volume 7079 of *Lecture Notes in Computer Science*, pages 234–251. Springer, 2011.